

Camila Duitama González

DATA SCIENTIST - BIOINFORMATICIAN ·

Paris, France

☎ (+33)749927193 | ✉ camiladuitama@gmail.com | 🏠 camiladuitama.github.io/gradfolio/ | 📷 CamilaDuitama | 🌐
camiladuitama | 🎓 Camila Duitama González



Summary

Python-focused software engineer with advanced expertise in backend development, data engineering, and algorithmic solutions. Proven ability to design, implement, and optimize complex software systems across research and commercial domains.

Skills

SQL & Database Management	Advanced SQL, MySQL, SQLite
Data Engineering & Analytics	Apache Spark, Pandas, NumPy, Data Pipeline Development
Machine Learning & Data Science	Scikit-learn, PyTorch, TensorFlow, Metaflow, Dask, Predictive Modeling
Programming & Development	Python, R, Bash, Docker, Git, Continuous Integration
Cloud & Infrastructure	AWS, High-Performance Computing, Slurm, Singularity, Cloud Workflows
Data Visualization & Reporting	Jupyter, R Markdown, LaTeX, Plotly, D3.js, Matplotlib
Bioinformatics Tools	Nextflow, Bioconda, Snakemake, GATK, Samtools
Languages	Spanish (Native), English (C2), French (B2), German (B1)

Experience

Sequence Bioinformatics Group, Institut Pasteur

Paris, France

POSTDOCTORAL RESEARCHER

January 2024 - Present

- Lead data engineering efforts for large-scale genetic research projects.
- Develop and optimized data pipelines for genotype-phenotype association studies.
- Implement machine learning models for complex biological data analysis.
- Mentor and supervise a team of computer scientists and programmers.

Max Planck Institute for Biology of Ageing

Cologne, Germany

RESEARCH ASSISTANT/PYTHON DEVELOPER

Jan-Dec 2018, Apr-Aug 2020

- Performed dimensionality reduction and multivariate analysis of heterogenous data types.
- Contributed to a Flask-based app for visualisation and analysis of biological data using Python and Docker.

Max Planck Institute of Biochemistry

Munich, Germany

INTERN

March 2017 - July 2017

- Conducted database curation, statistical analysis, visualization, and reporting for a multidisciplinary group of scientists.

Education

Sorbonne Université - Institut Pasteur

Paris, France

PH.D. INFORMATION AND COMMUNICATION SCIENCE AND TECHNOLOGY

October 2020 - January 2023

- Implemented and created machine learning algorithms for large-scale data classification and analysis.
- Developed predictive models using Python and various ML frameworks (Scikit-learn, Dask).
- Served as Teaching Assistant, demonstrating ability to convey complex technical concepts to diverse audiences.

Bonn University

Bonn, Germany

M.SC. IN LIFE SCIENCE INFORMATICS

Oct 2017 - Oct 2019

- Master thesis: Developed and validated a Bayesian model for risk prediction using R and Python.
- Gained experience in building predictive models and conducting statistical analyses on large datasets.

Universidad Nacional de Colombia

Medellín, Colombia

B.ENG. IN BIOLOGICAL ENGINEERING

Feb 2012 - Sept 2017

Open-Source Software Tools

Unitig Matrix Construction Utility

MUSET

C++

2025

- Software for constructing abundance unitig matrices from sequencing data.
- Open-source repository: github.com/CamilaDuitama/muset

Ancient DNA Decontamination Toolkit

AKMERBROOM

Python

2023

- Innovative Python tool for preprocessing and decontaminating ancient DNA samples.
- Open-source repository: github.com/CamilaDuitama/aKmerBroom

Microbial Source Tracking Tool

DECOM

C++

2023

- High-performance tool for similarity-based microbial source tracking.
- Developed for analyzing microbial composition in ancient oral samples.
- Open-source repository: github.com/CamilaDuitama/decOM

Presentations

Genome Informatics

Wellcome Genome Campus, UK

PRESENTER FOR "MUSET: SET OF UTILITIES FOR THE CONSTRUCTION OF ABUNDANCE UNITIG MATRICES FROM SEQUENCING

2024

DATA".

- Speaker in Session 2: Genome Assembly and Sequencing Algorithms

RECOMB-Seq 2023

Istanbul, Turkey

PRESENTER FOR "AKMERBROOM: ANCIENT ORAL DNA DECONTAMINATION USING BLOOM FILTERS ON K-MER SETS"

2023

- Delivered a proceedings talk on novel ancient DNA decontamination techniques.

Learning Meaningful Representations for Life (LMRL) Workshop, NeurIPS 2022

Virtual

ACCEPTED PAPER AND POSTER

2022

- Presented research poster in the proceedings track.

Publications

Journal Article

MICROBIOME

2023

- Duitama González, C., Vicedomini, R., Lemane, T., Rascovan, N., Richard, H., & Chikhi, R. (2023). decOM: similarity-based microbial source tracking of ancient oral samples using k-mer-based methods. *Microbiome*, 11(1), 243.

Journal Article

ISCIENCE

2023

- González, C. D., Rangavittal, S., Vicedomini, R., Chikhi, R., & Richard, H. (2023). aKmerBroom: Ancient oral DNA decontamination using Bloom filters on k-mer sets. *iScience*, 26(11).

Journal article

OXFORD BIOINFORMATICS

2025

- Vicedomini, R., Andreace, F., Dufresne, Y., Chikhi, R., & Duitama González, C. (under review). MUSET: Set of utilities for the construction of abundance unitig matrices from sequencing data.

References

Dr. Rayan Chikhi

Institut Pasteur, Paris, France

PI AT THE SEQUENCE BIOINFORMATICS GROUP

PhD/Postdoc supervisor

- Email: rchikhi@pasteur.fr
- LinkedIn: [linkedin.com/in/rayanchikhi](https://www.linkedin.com/in/rayanchikhi)

Dr. Hugues Richard

Robert Koch Institute / Sorbonne

University

RESEARCHER / ASSOCIATE PROF. (ON LEAVE)

PhD/Postdoc supervisor

- Email: richardh@rki.de
- LinkedIn: [linkedin.com/in/hugues-richard-36270185](https://www.linkedin.com/in/hugues-richard-36270185)

Dr. Ashar Ahmad

Grünenthal Group, Bonn, Germany

ASSOCIATE DIRECTOR, DATA SCIENCE & ADVANCED ANALYTICS

Master thesis supervisor

- Email: ashar799@gmail.com
- LinkedIn: [linkedin.com/in/ashar-ahmad-phd](https://www.linkedin.com/in/ashar-ahmad-phd)